A U-Net Based Discriminator for Generative Adversarial Networks

Abstract

Among the major remaining challenges for generative adversarial networks (GANs) is the capacity to synthesize globally and locally coherent images with object shapes and textures indistinguishable from real images. To target this issue we propose an alternative U-Net based discriminator architecture, borrowing the insights from the segmentation literature. The proposed U-Net based architecture allows to provide detailed per-pixel feedback to the generator while maintaining the global coherence of synthesized images, by providing the global image feedback as well. Empowered by the per-pixel response of the discriminator, we further propose a per-pixel consistency regularization technique based on the CutMix data augmentation, encouraging the U-Net discriminator to focus more on semantic and structural changes between real and fake images. This improves the U-Net discriminator training, further enhancing the quality of generated samples. The novel discriminator improves over the state of the art in terms of the standard distribution and image quality metrics, enabling the generator to synthesize images with varying structure, appearance and levels of detail, maintaining global and local realism. Compared to the BigGAN baseline, we achieve an average improvement of 2.7 FID points across FFHQ, CelebA, and the proposed COCO-Animals dataset.